# Prototype-Enhanced Recommendation with Synthetic Reviews

Group 5-2:
Wout Kooijman, Kai Liang, Weitao Luo, Erik Stammes, Ozzy Ülger, Vicky Foing

# Introduction

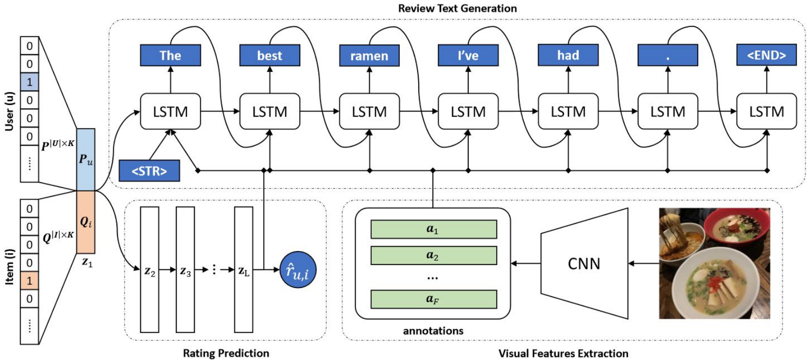
- Explainable recommendations for user satisfaction [3]
- Review generation for explainable recommendations
- Can we improve the existing state of the art to perform even better?

# Related Work

- Baselines (DeepCoNN [11], DER [12], NARRE[13])
- MTER [7]
- CTRL [8]
- Siamese LSTM networks [9]
- MRG [6]
- Prototype editing [5]
- Attention mechanism [10]

# Multimodal Review Generation (MRG)

- Model: A neural approach with two components:
    a. Rating prediction
    b. Review text generation
- Can use images to inform review text generation
- Performance (2019):
    a. Better rating prediction than MF models
    b. Better review text generation than LSTM models

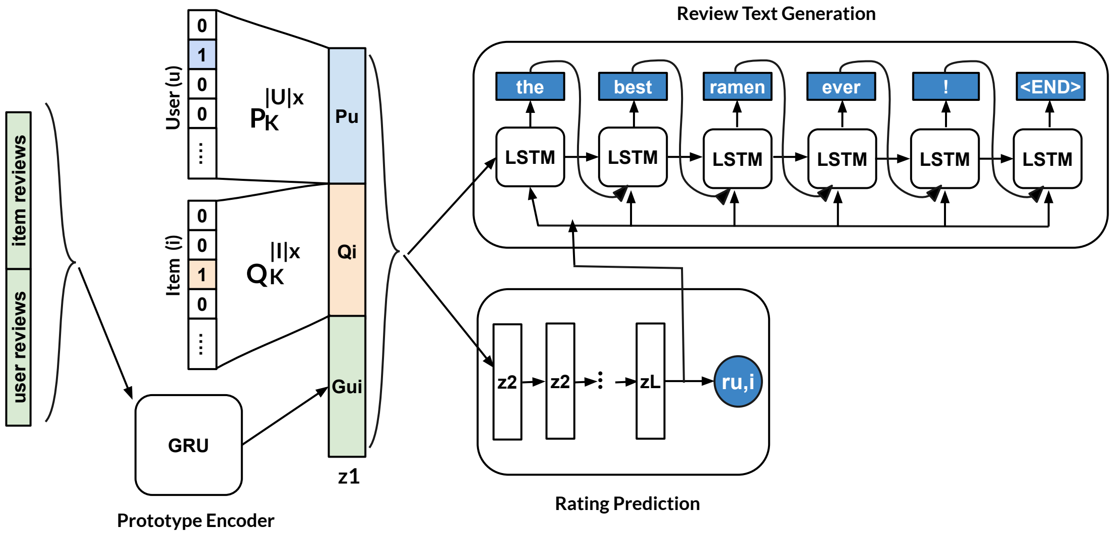


# Prototype editing

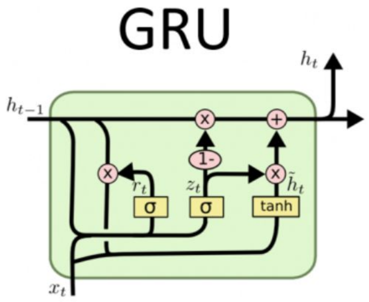
- Prototype editors used for
  - Generating sentences by editing random samples from training set [1]
  - Generating paragraphs by retrieving from training set and editing [2]
- Both use an RNN (LSTM) for encoding and decoding
- Attention mechanism is used to improve decoding

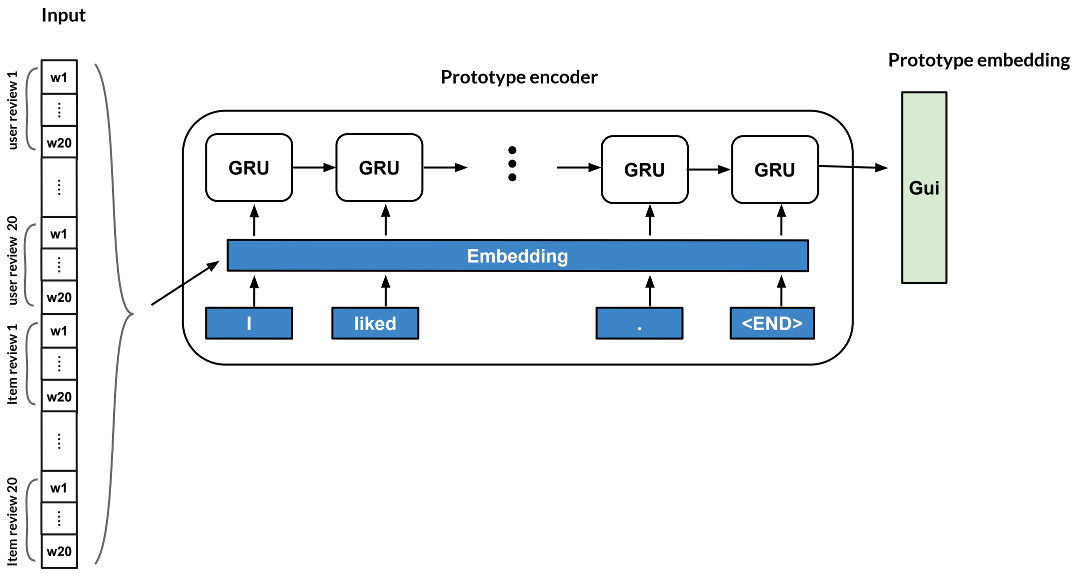
# Method

- Improve MRG architecture
  - Remove image component
  - Add prototype encoder



# Prototype encoder implementation



# Experiments

- Research questions
  - Does our model improve rating prediction?
  - Does our model improve review text generation?
  - What are the contributions of the the prototype encoder and attention mechanism?
- Ablation analysis
  - MRG without image component
  - Above, with prototypes
  - Above, with attention mechanism (in progress)

# Settings

- Yelp dataset split into training (80%), validation (10%), and testing (10%) sets
- Same hyperparameters used for baseline MRG and extended MRG
- GloVe pre-trained word embeddings [4] are used for training

| Learning rate | Dropout rate | Lambda (reg. term) | Number of epochs | Batch size | Word embedding dim. | Number of latent factors | LSTM hidden state dim. | Max. length of reviews |
|---|---|---|---|---|---|---|---|---|
| 3e-4 | 0.2 | 1e-4 | 20 | 64 | 200 | 256 | 256 | 20 |

# Evaluation

Measuring rating prediction:

- MAE: mean absolute error
- RMSE:  root mean squared error

Measuring semantic quality:

- BLEU (**precision**): How much does the generated review overlap with human review
- ROUGE (**recall**): How much does the human review overlap with generated review

# Results: Rating Prediction

| | MAE | RMSE |
|---|---|---|
| MRG* + Prototype Encoder | **0.786** | **1.024** |
| MRG* | 0.789 | 1.029 |

Lower errors for MRG + Prototype Encoder!

* MRG without image component

# Results: Review Text Generation (BLEU)

| | 1-gram | 2-gram | 3-gram | 4-gram |
|---|---|---|---|---|
| MRG* + Prototype Encoder | **37.12** | **20.47** | **15.21** | **13.22** |
| MRG* | 36.58 | 19.60 | 14.59 | 12.67 |

Higher BLEU scores for MRG + Prototype Encoder!

*MRG without image component

# Results: Review Text Generation (ROUGE)

| | 1-gram | | | 2-gram | | | L-gram (longest subsequence) | | |
|---|---|---|---|---|---|---|---|---|---|
| | Precision | Recall | F1 | Precision | Recall | F1 | Precision | Recall | F1 |
| MRG* + Prototype Encoder | 34.21 | **19.17** | **23.10** | **2.84** | **1.65** | **1.95** | **25.97** | **17.92** | **17.37** |
| MRG* | **34.24** | 18.61 | 22.65 | 2.38 | 1.35 | 1.60 | 24.36 | 17.45 | 16.76 |

Higher ROUGE F1 scores for MRG + Prototype Encoder!

*MRG without image component

# Insights

- **Prototype encoder** improves MRG
  - Lower MAE / RMSE = Better rating prediction
  - Higher BLEU scores (1-4)  = More of the generated review appears in the human review
  - Higher Rouge scores (1,2,L) = More of the human review appears in the generated review
- **Attention mechanism** is likely to improve MRG further (to be seen this week)

# Conclusion

- We tried many other different approaches to outperform baselines
  - CTRL
  - Prototype editor from Guu et al [1]
  - Prototype editor from Hashimoto et al [2]
  - Siamese LSTMs
- So far, MRG + prototype encoder is the best solution
- Still experimenting with attention mechanism
- **Future directions**: Prototype decoder, Siamese LSTMs

# References

**[1] Guu, K., Hashimoto, T. B., Oren, Y., & Liang, P. (2018).** Generating sentences by editing prototypes. Transactions of the Association for Computational Linguistics, 6, 437-450.
**[2] Hashimoto, T. B., Guu, K., Oren, Y., & Liang, P. S. (2018).** A retrieve-and-edit framework for predicting structured outputs. In Advances in Neural Information Processing Systems (pp. 10052-10062).
**[3] Zhang, Y., & Chen, X. (2018).** Explainable recommendation: A survey and new perspectives. arXiv preprint arXiv:1804.11192.
**[4] Pennington, J., Socher, R., & Manning, C. (2014, October).** Glove: Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP) (pp. 1532-1543).
**[5] Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014).** Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078.
**[6] Truong, Quoc-Tuan, and Hady Lauw (2019).** "Multimodal Review Generation for Recommender Systems." The World Wide Web Conference. ACM.
**[7] Wang, Nan, et al (2018).** "Explainable recommendation via multi-task learning in opinionated text data." The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval. ACM.
**[8] Keskar, N. S., McCann, B., Varshney, L. R., Xiong, C., & Socher, R. (2019).** Ctrl: A conditional transformer language model for controllable generation. arXiv preprint arXiv:1909.05858.
**[9] Mueller, J., & Thyagarajan, A. (2016, March).** Siamese recurrent architectures for learning sentence similarity. In Thirtieth AAAI Conference on Artificial Intelligence.
**[10] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017).** Attention is all you need. In Advances in neural information processing systems (pp. 5998-6008).

# References

**[11] Zheng, Lei, Vahid Noroozi, and Philip S. Yu.** "Joint deep modeling of users and items using reviews for recommendation." Proceedings of the Tenth ACM International Conference on Web Search and Data Mining. ACM, 2017.
**[12] Chen, Xu, Yongfeng Zhang, and Zheng Qin.** "Dynamic Explainable Recommendation based on Neural Attentive Models." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 33. 2019.
**[13] Chen, Chong, et al.** "Neural attentional rating regression with review-level explanations." Proceedings of the 2018 World Wide Web Conference.

# Questions?

Thank you!